

**Section III:**  
**AMENDMENT UNDER 37 CFR §1.121 to the**  
**DRAWINGS**

No amendments or changes to the Drawings are proposed.

**Section IV:**  
**AMENDMENT UNDER 37 CFR §1.121**  
**REMARKS**

**Rejections under 35 U.S.C. §102(e)**

We appreciate the reconsideration of the rejections of Claims 1 - 5, 10 - 18 under 35 U.S.C. §102(e) over Burdick.

**Unclear Finality of Rejections**

We noticed that the PAIR status and the Summary of the Office Action indicate the present rejections are non-final, but the last paragraph of the Office Action indicates it is final. The Examiner has indicated by telephone today (Feb. 13, 2008) that the last paragraph is a typographical error, and that the action is meant to be non-final.

We appreciate the Examiner's assistance with this matter, and we appreciate non-finality of the Office Action.

**Rejections under 35 U.S.C. §103(a)**

Regarding the rejections of claims 1 - 5 and 10 - 18 under 35 U.S.C. §103(a) over Burdick in view of US pre-grant patent application publication 2006/0161814 to Wocke, we appreciate the opportunity to explain why we believe our claims are patentable.

Claim 1. We agree with the Examiner's finding that Burdick fails to teach our claim elements of:

*"generating a set of cleaning attributes for each cleaned data record in a complete set of cleaned data records, said cleaning attributes reflecting which fields of each record have been modified by a cleaning operation";*

*"receiving a data feature identified by a data mining process for a subset of said complete set of cleaned data records";*

*"determining a degree of correlation of said data feature to the modified fields of said subset of cleaned data records according to said cleaning attributes" . . .*

Several paragraphs of Wocke's disclosure were cited in the reasons for the rejection. We are not certain which features within those paragraphs were believed to teach our *data cleaning flags*. For example, ¶0023 lists an *attribute matrix*, a *range filter*, a *zooming function*, a *labeling engine*, and a *search function*, and ¶0039 describes a *knowledge filter*. For the purposes of this reply, we assume that it was the *attribute matrix* which the Examiner believed taught our *cleaning attributes*. We believe that ¶0111 was relied upon for its disclosure of binary or Boolean flags in general. We believe that ¶0222 - 0225 were relied upon for their disclosure of an *afterGrow* flag, whereas Wocke's *MatchList* and *randomization* do not appear to have direct bearing on changing actual data values. If these assumptions are incorrect, we would appreciate clarification in the next Office Action.

We ask the Examiner to reconsider our claim phrase as a whole:

*"generating a set of cleaning attributes for each cleaned data record in a complete set of cleaned data records, said cleaning attributes reflecting which fields of each record have been modified by a cleaning operation".*

We believe that one of ordinary skill in the art, using the plain meanings of these terms coupled with the information from our disclosure, would interpret this phrase to mean that there is a cleaning attribute *for each field in each data record*. Normally, when one refers to a *data record* having *fields*, it is meant and understood that each *record* has multiple *fields*. Thus, we believe that our claim language "*cleaning attributes for each cleaned data record*" explicitly states that there is a set of cleaning attributes for each record. And, we believe that our claim language "*said cleaning attributes reflecting which fields of each record have been modified*" further specifies that *each field* has its own cleaning attribute. We have provided examples of this arrangement and relationship in our disclosure, such as that shown in ¶0069, and in Figs. 5a and 5b. If the Examiner disagrees that this language would be interpreted in this manner by one of ordinary skill in the art, we would appreciate any suggestions for re-phrasing that the Examiner can offer.

We believe that one of ordinary skill in the art applying plain meanings of the terms used in Wocke's disclosure would read their *afterGrow* attribute to be a single attribute associated with an entire set of data records and all of the fields within the data records. Their ¶0227 describes their *afterGrow* flag as indicating whether or not their *map* was changed through their

growing algorithm. Their ¶0041 and 0043 define their "map" as the output data set of records (e.g. multiple records). Their *growing algorithm* is a method for *interpolating* new values and *inserting* them into the data records ¶0012, presumably at fields within the records. We believe Wocke is using the term "attribute" to also refer to "fields" within records, as this is sometimes done in database vernacular (see their ¶0065, for example).

As such, we believe that their *afterGrow* flag is a singular flag which is set true if any data values have been interpolated and inserted anywhere in the entirety of the set of data records and their fields.

While this is useful to know if an entire set of data records should be suspect because somewhere within it values were changed or inserted (e.g. "cleansed"), we do not believe this is same as or as useful as adding a data cleansing flag *for each field within each record*. With our data cleansing attributes, the processes which subsequently use the cleansed data records can determine *exactly* which fields have been modified, not just which records or even which sets of records.

Please note also that the *data feature*, such as a cluster, received in our claims applies to a subset of the complete set of cleaned data records. We believe that one of ordinary skill in the art would read this to mean that, for example, a cluster was found *within the full set of records*. So, with our "per field" cleansing attributes, it could then be determined if the data feature (e.g. the cluster in this case) is "suspect" because the subset of fields includes a cleansed field. This determination is not possible with the combination of Burdick and Wocke because Wocke's *afterGrow* flag applies to the entirety of the data set, not just to specific fields, and therefore all *data features* regarding a *subset of the records* would be *suspect*.

We do not believe that it would have been obvious to one of ordinary skill in the art to further modify Wocke, or Burdick, to record a cleansing attribute *for each field in each record* of a set of data records. If one uses Wocke and Burdick as indicators of the ordinary skill level in the art at the time of filing our application, then it is noticeable that neither Burdick or Wocke recognize the problem of relying upon data features in subsets of data records which include cleansed data fields. Wocke recognizes a need to generally know if a data set has been grown, but does not recognize the need to further identify *which fields* within that data set were grown (or otherwise modified) so that any data features including those modified fields could be considered *suspect* (e.g. given less or no weight in decision making).

So, we respectfully request reconsideration of the teachings of Burdick in view of Wocke and review of our claim language. We believe that the claims as written set forth steps, elements, and limitations not taught by Burdick or Wocke, which would not have been obvious to one of ordinary skill in the art at the time of our invention.

Claim 2. We respectfully disagree that Wocke's ¶0045 "*symbol meaning*" and ¶0121-0124 disclose "*bit-mapped*" cleansing attributes *for each field in each record*. ¶0045 contains a list of multi-bit values, thus it is not bit-mapped. And, ¶0124 discloses their single *afterGrow* flag for the entire "map" (e.g. for the entire output set of data records), and other multi-bit values (e.g. maximum number of iterations, minimum quantization threshold, current step number, etc.), which is not a bit-map.

Claim 3. We disagree that Wocke discloses appending or prepending our bit-mapped set of cleaning attributes to *each* record within the data set. We believe Wocke only discloses one *afterGrow* flag for the entire set of records, thus it is not appended to *each record*, but instead is associated with the entire set of records.

Claim 4. We agree that Wocke discloses receiving or finding a *cluster* as a *data feature*. However, Claim 4 depends from Claim 1, which recites steps, elements and features (e.g. one cleansing attribute *per field in each record*), which are not taught or suggested by Burdick or Wocke.

Claim 5. Claim 5 depends from Claim 1, which recites steps, elements and features (e.g. one cleansing attribute *per field in each record*), which are not taught or suggested by Burdick or Wocke.

Claims 10 - 14. We believe that Claims 10 - 14 recite patentably distinctive steps, elements, and features over Burdick in view of Wocke as set forth in our arguments regarding Claims 1 - 5, respectively.

Claims 15 - 18. We believe that Claims 15 - 18 recite patentably distinctive steps,


elements, and features over Burdick in view of Wocke as set forth in our arguments regarding Claims 1 - 5, respectively.

**Request for Indication of Allowable Subject Matter**

We believe we have responded to all grounds of rejection and objection, but if the Examiner disagrees, we would appreciate the opportunity to supplement our reply.

We believe the present amendment places the claims in condition for allowance. If, for any reason, it is believed that the claims are not in a condition for allowance, we respectfully request constructive recommendations per MPEP 707.07(j) II which would place the claims in condition for allowance without need for further proceedings. We will respond promptly to any Examiner-initiated interviews or to consider any proposed examiner amendments.

Respectfully,

A handwritten signature in black ink that reads "Robert Frantz". The signature is written in a cursive, flowing style with a large, stylized "R" and "F".

Robert H. Frantz, Reg. No. 42,553  
Tel: (405) 812-5613  
Franklin Gray Patents, LLC